

CLAIM AMENDMENTS

1. (Currently amended) A method for analyzing a plurality of sets of values associated with a plurality of genes~~genes or proteins~~ to identify genes~~genes or proteins~~ whose associated values differ by an amount of statistical significance among the sets, ~~wherein each of the sets of associated values of the genes is obtained from one of a number of data sources,~~ said associated values comprising levels of ~~gene~~ gene or protein expression or levels of proteins encoded by genes, wherein the method comprises:

providing for each of the plurality of genes or proteins a parameter that contains information concerning differences in the associated values of that gene or protein among the sets;

adjusting the parameters of the plurality of genes or proteins so that the parameters are substantially independent of scatter values or average associated values of the genes~~genes or proteins~~ over the sets;

deriving an observed value and an expected value of the adjusted parameter for each gene~~gene or protein~~ from the sets of associated values; and

comparing the observed and expected values of the parameter to identify genes~~genes or proteins~~ whose associated values differ by an amount of statistical significance among the sets.

2. (Currently amended) The method of claim 1, wherein said adjusting includes:

dividing the scatter values or average associated values of the genes~~genes or proteins~~ into subsets each having a similar range of values, and calculating the standard deviation of each of the parameters within each subset;

altering the parameters until a coefficient of variation of the standard deviations of the parameters among the subsets is minimized.

3. (Currently amended) The method of claim 1, further comprising obtaining said sets of associated values from multiple measurements of the plurality of genes~~genes or proteins~~, or values derived therefrom.

4. (Original) The method of claim 1, wherein said sets of associated values represent gene expression or number of gene copies or levels of protein encoded by the genes.

5. (Original) The method of claim 1, wherein said sets of associated values include calculated or predicted values.

6. (Currently amended) The method of claim 1, wherein said providing includes calculating a difference value between an associated value of each ~~gene-gene or protein~~ in a first of the sets or a value derived therefrom and an associated value of that ~~gene-gene or protein~~ in a second of the sets or a value derived therefrom; wherein the parameter is a function of the difference value of that gene.

7. (Currently amended) The method of claim 6, wherein said providing further includes:

generating for each of the plurality of ~~genes~~genes or proteins a scatter value that quantifies variation in the associated values of that ~~gene-gene or protein~~ within the first and second sets; and wherein said parameter is a function of the scatter value and of the difference value, said parameter defining a relative difference value of that gene.

8. (Currently amended) The method of claim 7, wherein said generating employs the following equation:

$$s(i) = (\{1/a\} \{ \sum_m [x_m(i) - \underline{x}_I(i)]^2 + \sum_n [x_n(i) - \underline{x}_U(i)]^2 \})^{1/2}$$

where gene (i) or protein (i) has associated values $x_I(i)$ and $x_U(i)$ in Ith and Uth states respectively in the first and second sets of associated values, I and U being positive integers; \sum_m and \sum_n are sums over associated values of gene (i) or protein (i) in states I in the first set and in states U in the second set respectively, where $s(i)$ is the scatter value of gene (i) or protein (i), and a is a constant.

9. (Original) The method of claim 8, wherein said calculating calculates the parameter $d(i)$ from the following equation:

$$d(i) = [\underline{x}_I(i) - \underline{x}_U(i)]/[s(i) + s_o]$$

where s_0 is a constant, and $\bar{x}_I(i)$ and $\bar{x}_U(i)$ are the average values of $x_I(i)$ and $x_U(i)$ respectively in the first and second sets of associated values.

10. (Currently amended) The method of claim 9, further comprising:

dividing the scatter values or average associated values of the genesgenes or proteins into subsets each having a similar range of values, and calculating the standard deviation of each of the parameters within each subset; and

altering value of s_0 until a coefficient of variation of the standard deviations of the parameters among the subsets is minimized.

11. (Currently amended) The method of claim 1, wherein said associated values of the genesgenes or proteins are correlated with another variable so that each of said associated values has a corresponding value of the variable, and wherein the parameter is provided using a Pearson correlation coefficient related to a weighted difference between each of the associated values and an average associated value, the variance of the associated values and the variance of the variable, said difference weighted by deviation of the corresponding value of the variable of such associated value from its average value.

12. (Original) The method of claim 11, wherein said variable is continuous.

13. (Original) The method of claim 12, wherein said variable is time.

14. (Currently amended) The method of claim 11, wherein the parameter is selected using the Pearson correlation coefficient and a quantity s_0 that has a value adjusted as follows:

dividing the scatter values or average associated values of the genesgenes or proteins into subsets each having a similar range of values, and calculating the standard deviation of each of the parameters within each subset; and

altering value of s_0 until a coefficient of variation of the standard deviations of the parameters among the subsets is minimized.

15. (Currently amended) The method of claim 11, the number of sets of associated values being k, k being a positive integer, wherein said Pearson correlation coefficient $r(i)$ is given by:

$$r(i) = \sum_k [(x_k(i) - \underline{x}(i))[(y_k - \underline{y})] / \sqrt{\sum_k (x_k(i) - \underline{x}(i))^2 \sum_k (y_k - \underline{y})^2}$$

where $x_k(i)$ is the associated value of gene (i) or protein (i) in the kth set of associated values, $\underline{x}(i)$ the average of the associated values of gene (i) or protein (i) in all the sets, y_k the value of the variable corresponding to $x_k(i)$, \underline{y} the average value of y_k in all the sets, and \sum_k is a sum over all values of k.

16. (Original) The method of claim 1, wherein the associated values in each set are classified into two or more subsets with values in each subset having a correlation with one another, and wherein the parameter is selected using a quantity related to variances between the associated values in the subsets of the sets and the variances of the associated values within each subset of the sets.

17. (Original) The method of claim 16, wherein the quantity relates to the sum of variances between the associated values in the subsets of the sets and the sum of variances of the associated values within each subset of the sets.

18. (Currently amended) The method of claim 17, wherein the parameter is selected using the Fisher discriminant and a quantity s_0 having a value which has been adjusted as follows:

dividing the scatter values or average associated values of the ~~genes~~genes or proteins into subsets each having a similar range of values, and calculating the standard deviation of each of the parameters within each subset; and

altering value of s_0 until a coefficient of variation of the standard deviations of the parameters among the subsets is minimized.

19. (Currently amended) The method of claim 18, wherein the number of subsets of associated values of such set being k, k being a positive integer, and the Fisher discriminant $F(i)$ is given by:

$$F(i) = \sum_k n_k [x_k(i) - \bar{x}(i)]^2 / \sum_k \sum_j [x_j(i) - \bar{x}_k(i)]^2$$

where $x_k(i)$ is an associated value of gene (i) or protein (i) in the kth subset of associated values, $\bar{x}_k(i)$ the average of the associated values of gene (i) or protein (i) in the kth subset, $\bar{x}(i)$ the average value of the associated values of gene (i) or protein (i) in all of the subsets, n_k the number of associated values in the kth set, \sum_j a sum over all the associated values of gene (i) or protein (i) in the kth subset, and \sum_k a sum of the associated values of gene (i) or protein (i) over all of the subsets.

20. (Currently amended) The method of claim 1, the sets of associated values referred to as original sets, wherein said deriving includes deriving said expected value by:

permuting, for each of the plurality of genesgenes or proteins, the associated values for such gene-gene or protein in the original sets to arrive at a number of different permutations;

classifying the associated values in each permutation of each gene-gene or protein into corresponding permuted sets that are different from the original sets; and

supplying for each permutation a parameter value of each of the genesgenes or proteins derived from an associated value of such gene-gene or protein in each of the corresponding permuted sets for such permutation or values derived therefrom.

21. (Currently amended) The method of claim 20, wherein said associated values of the genesgenes or proteins are correlated with another variable so that each of said associated values has an associated value of the variable, wherein the permuting permutes the associated values so that at least each of some of the associated values has a different associated variable.

22. (Original) The method of claim 21, wherein the associated values are classified into two or more subsets with values in each subset having a correlation with one another, wherein the permuting permutes the associated values so that at least each of some of the associated values is in a subset different from the subset it is classified into.

23-27. (Cancelled)

28. (Currently amended) A method for analyzing a plurality of original sets of values associated with a plurality of ~~genes~~genes or proteins to identify ~~genes~~genes or proteins whose associated values differ by an amount of statistical significance among the sets, ~~wherein each of the sets of associated values of the genes is obtained from one of a number of data sources,~~ wherein the method comprises:

calculating for each ~~gene~~gene or protein a value for a statistical parameter indicating differences between associated values of such ~~gene~~gene or protein among the original sets;

ranking the values of the parameter of the ~~genes~~genes or proteins;

providing an expected value of such parameter for each rank, wherein said providing includes permuting the associated values in the original sets to arrive at sets different from the original sets for each permutation, deriving a value of such parameter for each permutation, and ranking such values; and

comparing the calculated and expected values for the parameter of the same rank to identify ~~genes~~genes or proteins whose associated values differ by an amount of statistical significance among the sets.

29. (Currently amended) The method of claim 28, wherein said providing comprises:

for each permutation, deriving a value of the parameter for each ~~gene~~gene or protein and ranking the ~~genes~~genes or proteins by their associated parameter values; and

determining the expected value of such parameter for each rank by computing an average value of the parameter of all the permutations having such rank.

30. (Currently amended) The method of claim 29, wherein said comparing comprises identifying a ~~gene~~gene or protein as one whose associated values differ by an amount of statistical significance among the sets when the difference for such ~~gene~~gene or protein between the calculated value of the parameter of a rank and the expected value of such parameter of the same rank exceeds a threshold.

31-32. (Cancelled)

33. (Original) The method of claim 28, wherein the sets of associated values in each permutation contains approximately an equal number of associated values from each of the original sets of associated values.

34-43. (Cancelled)

44. (Currently amended) A computer readable storage device embodying a program of instructions executable by a computer to perform a method for analyzing a plurality of sets of values associated with a plurality of genes~~genes~~ or proteins to identify genes~~genes~~ or proteins whose associated values differ by an amount of statistical significance among the sets, ~~wherein each of the sets of associated values of the genes is obtained from one of a number of data sources,~~ wherein the method comprises:

providing for each of the plurality of genes~~genes~~ or proteins a parameter that contains information concerning differences in the associated values of that ~~gene-gene or protein~~ among the sets;

adjusting the parameters of the plurality of genes~~genes~~ or proteins so that the parameters are substantially independent of scatter values or average associated values of the genes~~genes~~ or proteins over the sets;

deriving an observed value and an expected value of the adjusted parameter for each gene-gene or protein from the sets of associated values; and

comparing the observed and expected values of the parameter to identify genes~~genes~~ or proteins whose associated values differ by an amount of statistical significance among the sets.

45. (Cancelled)

46. (Currently amended) A computer readable storage device embodying a program of instructions executable by a computer to perform a method for analyzing a plurality of original sets of values associated with a plurality of genes~~genes~~ or proteins to identify genes~~genes~~ or proteins whose associated values differ by an amount of statistical

significance among the sets, ~~wherein each of the sets of associated values of the genes is obtained from one of a number of data sources,~~ wherein the method comprises:

calculating for each ~~gene-gene or protein~~ a value for a statistical parameter indicating differences between associated values of such ~~gene-gene or protein~~ among the original sets;

ranking the values of the parameter of the ~~genes~~genes or proteins;

providing an expected value of such parameter for each rank, wherein said providing includes permuting the associated values in the original sets to arrive at sets different from the original sets for each permutation, deriving a value of such parameter for each permutation, and ranking such values; and

comparing the calculated and expected values for the parameter of the same rank to identify ~~genes~~genes or proteins whose associated values differ by an amount of statistical significance among the sets.

47-57. (Cancelled)

58. (Currently amended) A computer system for analyzing a plurality of sets of values associated with a plurality of ~~genes~~genes or proteins to identify ~~genes~~genes or proteins whose associated values differ by an amount of statistical significance among the sets, ~~wherein each of the sets of associated values of the genes is obtained from one of a number of data sources,~~ wherein the system comprises:

one or more computers;

one or more computer programs running on the computer(s), performing the following:

providing for each of the plurality of ~~genes~~genes or proteins a parameter that contains information concerning differences in the associated values of that ~~gene-gene or protein~~ among the sets;

adjusting the parameters of the plurality of ~~genes~~genes or proteins so that the parameters are substantially independent of scatter values or average associated values of the ~~genes~~genes or proteins over the sets;

deriving an observed value and an expected value of the adjusted parameter for each ~~gene~~gene or protein from the sets of associated values; and

comparing the observed and expected values of the parameter to identify ~~genes~~genes or proteins whose associated values differ by an amount of statistical significance among the sets.

59. (Cancelled)

60. (Currently amended) A computer system for analyzing a plurality of original sets of values associated with a plurality of ~~genes~~genes or proteins to identify ~~genes~~genes or proteins whose associated values differ by an amount of statistical significance among the sets, ~~wherein each of the sets of associated values of the genes is obtained from one of a number of data sources,~~ wherein the system comprises:

one or more computers;

one or more computer programs running on the computer(s), performing the following:

calculating for each ~~gene~~gene or protein a value for a statistical parameter indicating differences between associated values of such ~~gene~~gene or protein among the original sets;

ranking the values of the parameter of the ~~genes~~genes or proteins;

providing an expected value of such parameter for each rank, wherein said providing includes permuting the associated values in the original sets to arrive at sets different from the original sets for each permutation, deriving a value of such parameter for each permutation, and ranking such values; and

comparing the calculated and expected values for the parameter of the same rank to identify ~~genes~~genes or proteins whose associated values differ by an amount of statistical significance among the sets.

61-64. (Cancelled)